Tests in Censored Models when the Structural Parameters Are Not Identified

Leandro M. Magnusson

Department of Economics

Tulane University

lmagnuss@tulane.edu

December 2, 2008

Abstract

This paper presents tests for the structural parameters of a censored regression model with endogenous explanatory variables. These tests have the correct size even when the identification condition for the structural parameter is invalid. My approach starts from the estimation of the unrestricted parameters, which does not depend on the identification of the structural parameter. Next, I set up the optimal minimum distance objective function, from where I derive the tests. The proposed robust tests are implemented in many statistical software packages since they demand only the 'Tobit' and the 'ordinary least squares' estimation functions. By simulating their power curves, I compare the robust to the Wald and the likelihood ratio tests. A case of the labor supply of married women illustrates the use of the robust tests for the construction of confidence intervals.

JEL Classification: C12, C34.

Keywords: Endogenous Tobit, weak instruments, minimum distance estimation, female labor supply.

1 Introduction

The purpose of this paper is to present tests for the structural parameters of censored models which have the correct size under the null hypothesis even when those parameters are not identified. These tests depart from the minimum distance objective function and can be performed in many statistical software packages.

In nonlinear models, general identification conditions for the structural parameters are hard to obtain. However, a necessary global identification condition is that the expected value of the Jacobian of the objective function under the true distribution must be a full rank matrix (Newey and McFadden (1994)). The lack of identification misguides the usual asymptotic theory behind point estimation and hypothesis testing (see Staiger and Stock (1997) and Stock and Wright (2000)).

New tests have been developed to overcome the deficiencies of the Wald, Lagrange multiplier and likelihood ratio tests when the identification condition fails. In the case of linear simultaneous equation model, the pioneering test is the AR-test (Anderson and Rubin (1949)). Kleibergen (2002) proposes a Lagrange multiplier test, also known as the K-test, based on the asymptotic independence between the empirical moment restriction and its Jacobian under the null hypothesis. This principle comes from the partition of an invariant sufficient statistic into two independent ones. Then, tests can be performed conditioning one statistic into the other. Using this principle Moreira (2003) derives the conditional Wald (CW) and the conditional likelihood ratio (CLR) tests which are not pivotal. Departing from the objective function of the continuous updating estimator (CUE), Kleibergen (2005) extends the K- and the CLR-tests to the generalized method of moments framework.

However, for models in which the structural parameters are not separable from the others, the K-test demands the identification and the consistent estimation of untested parameters under the null hypothesis. It also requires the estimation of the covariance between the empirical moments and the Jacobian. Often, the estimation of untested parameters and the covariance matrix between moments and the Jacobian are computationally intensive.

I derive weak instruments robust tests for censored models departing from the minimum distance objective function. This approach avoids the estimation of untested parameters and covariance matrix between moments and Jacobian under the null hypothesis. Moreover, they can be implementable using regular statistical software such as *Stata* and *R*. Those robust tests are modifications of existing ones, so I use the subscript M to denote them.

I first define the unrestricted model and the linear restriction mapping between auxiliary and structural parameters. In the unrestricted model, the auxiliary parameters are well-identified independent of the presence of weak instruments. The auxiliary parameters can be estimated either by the two-stage conditional maximum likelihood method as proposed by Smith and Blundell (1986), or any other consistent estimators such as the symmetrically censored least square (Powell (1986)) and the winsorized mean estimator (Lee (1995)). Simple linear restrictions on the unrestricted parameters are enough to obtain the minimum distance objective function for the structural parameter. Robust tests are derived from the minimum distance objective function following the same lines as Kleibergen (2005). The minimum distance approach allows the extension of the weak instrument robust tests to other classes of limited dependent variable models such as endogenous probit and endogenous ordered probit (see Magnusson (2006)).

In the next section I present the censored model with endogenous explanatory variables and the assumptions behind it. I also discuss the failure of identification and the weak instruments asymptotics for that model. The third section deals with the derivation of weak instruments robust tests using the minimum distance approach. The fourth section presents simulations of the rejection probability curves in order to compare the performance between the proposed tests and the Wald and likelihood ratio tests. In the fifth section I use the weak instruments robust tests to build confidence intervals for the structural parameter of a female labor supply model. The sixth section summarizes and concludes the paper. Proofs, mathematical passages and data description are in the appendices.

2 The Censored Model with Endogenous Explanatory Variables

2.1 Model and Identification

The censored model with endogenous explanatory variables, also known as the endogenous Tobit, is first addressed by economic literature in the seventies (see Amemiya (1979) and Lee (1981)). It departs from the following structural latent linear simultaneous equation model:

$$\begin{cases} Y_t^* = X_t\beta + U_t \\ X_t = Z_t\Pi_z + V_t \end{cases}$$

where Y_t^* and U_t are scalars, X_t and V_t are $1 \times m$ vectors of endogenous variables and residuals, Z_t is $1 \times k$ vector of excluded instruments. Y_t^* is observed only if $Y_t^* > 0$. Included exogenous variables are omitted from the model for the ease of exposition only. The simultaneous equation system with one censored endogenous variable is defined as:

$$\begin{cases} Y_t = \max\{0, X_t\beta + U_t\} \\ X_t = Z_t\Pi_z + V_t \\ D_t = 1(Y_t^* > 0) \end{cases} \tag{2.1}$$

where $1(\cdot)$ is a binary indicator function which assumes the value 1 if $Y_t^* > 0$ and 0 otherwise. Residuals follow an independent multivariate normal distribution.[1]

Assumption 1. *Let $\{U_t, V_t\}_{t=1}^{T}$ be a sequence of independent random variables. Each pair $\{U_t, V_t\}$ follows a multivariate normal distribution conditional on Z_t, i.e,*

$$(U_t, V_t)|Z_t \sim \mathrm{N}\left(0, \begin{bmatrix} \sigma_{\mathrm{u}}^2 & \Sigma_{\mathrm{uv}} \\ \Sigma_{\mathrm{vu}} & \Sigma_{\mathrm{vv}} \end{bmatrix}\right)$$

There are several ways to estimate the parameters of the above limited information model under assumption 1. Some examples include the maximum likelihood estimator (MLE), the Amemiya generalized least squares (AGLS - Amemiya (1979)), the two-stage conditional maximum likelihood (TSCML - Smith and Blundell (1986)) and the Newey conditional generalized least squares (CGLS - Newey (1987)).

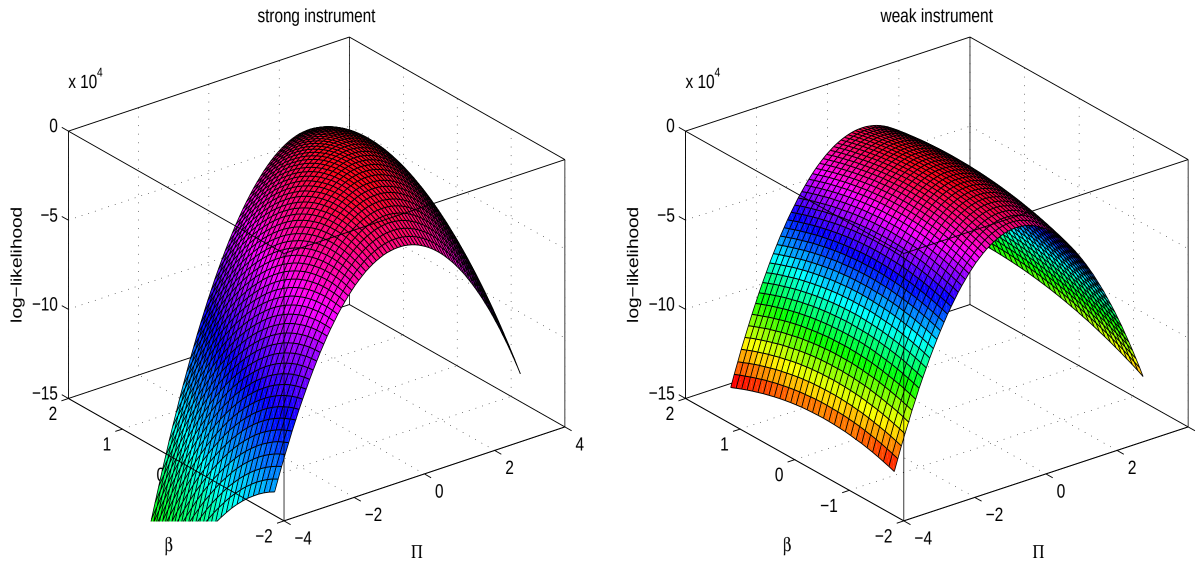
The regular identification condition for the structural parameter β requires that Π_z is full column ranked, i.e., the instruments should be correlated with the endogenous variables. If there exists a non-full rank matrix in a "small" neighborhood of Π_z or if Π_z is not itself a full rank matrix, then the exogenous variables are labeled as weak instruments. In their presence, the small sample distribution of a estimator is different from the asymptotic approximation. Additionally, statistical inference based on the classical tests (Wald, Lagrange multiplier (LM) and likelihood ratio (LR)) depends on the presence of nuisance parameters (Stock and Wright (2000)) and so is invalid.

Figure 1 illustrates the log-likelihood functions when the instruments are strong and weak. In this example there is only one instrumental variable Z_t, which follows a standard normal distribution. I set $\Pi_z = 1$ and $\Pi_z = 0.1$ in order to mimic, respectively, strong and weak instruments. The residuals U_t and V_t are joint-normally distributed with $\sigma_{\mathrm{u}}^2 = \Sigma_{\mathrm{vv}} = 1$ and $\Sigma_{\mathrm{uv}} = 0.5$. The log-likelihood functions are evaluated assuming that the covariance terms are known.

In the case of the strong instrument, the log-likelihood is globally concave and is uniquely maximized. When the instrument is weak the log-likelihood resembles a quasiconcave function. The smoothness along the line where $\Pi_z = 0$ indicates the lack of global identification of β.

[1]This assumption can be relaxed.

Fig. 1: Endogenous Tobit log-likelihood functions with strong and weak instrument.



Staiger and Stock (1997) model Π_z as local to zero in order to describe weak instruments asymptotics. I adopt the same assumption which is reproduced below as a definition.

Definition 1. *Let C be a full rank matrix. Π_z has the following asymptotic behavior in case of strong, weak and irrelevant instruments, respectively:*

i) $\Pi_z = C$,

ii) $\Pi_z = \Pi_T = \frac{C}{\sqrt{T}}$,

iii) $\Pi_z = 0$.

2.2 Likelihood and Score for the Structural Model

From the properties of the multivariate normal distribution, we have:

$$U_t = V_t\alpha + \varepsilon_t, \quad \varepsilon_t|V_t,Z_t \sim \mathcal{N}(0, \sigma_\varepsilon^2)$$

where $\alpha = \Sigma_{\mathrm{vv}}^{-1}\Sigma_{\mathrm{vu}}$, $\sigma_\varepsilon^2 = \sigma_{\mathrm{u}}^2(1-\rho'\rho)$ and $\rho = \frac{\Sigma_{\mathrm{vv}}^{-1/2}\Sigma_{\mathrm{vu}}}{\sigma_{\mathrm{u}}}$. The conditional structural model is obtained by substituting the above relation into equation (2.1).

$$\begin{cases} Y_t = \max\{0, X_t\beta + V_t\alpha + \varepsilon_t\} \\ X_t = Z_t\Pi + V_t \\ D_t = 1(Y_t^* > 0) \end{cases} \tag{2.2}$$

The limited information density function derived from (2.2) can be decomposed between marginal and conditional distributions, since ε_t is conditionally independent of X_t. The log-likelihood function for the endogenous Tobit after concentrating Σ_{vv} out is:

$$\ell_T(\beta, \alpha, \sigma_\varepsilon^2, \Pi_z, \Sigma_{\text{vv}}) = \sum_{t=1}^{T} \ell_{y|(x,z)}(y_t|x_t, z_t; \beta, \alpha, \sigma_\varepsilon^2, \Pi_z) + \sum_{t=1}^{T} \ell^c_{x|z}(x_t|z_t; \Pi_z)$$

where $\ell_{y|(x,z)}$ is the log-likelihood of the Tobit model with latent mean $x_t\beta + v_t\alpha$ and variance σ_ε^2 and $\ell^c_{x|z}$ is the concentrated log-likelihood of the multivariate normal density.

Setting moment restrictions which are valid under the null hypothesis is the starting point for testing the value of the structural parameter β. In the maximum likelihood set up, the score functions provide the natural moments. Before presenting them, let me introduce more notations. The following vector

$$\eta = \left[\alpha' \quad \sigma_\varepsilon^2 \quad \text{vec}(\Pi_z)'\right]',$$

is defined as the vector of parameters that are not being tested. The pseudo-residuals are farther defined as follows:[2]

$$e_t^{(1)}(\beta, \eta) = d_t \left(\frac{y_t - w_t\delta}{\sigma_\varepsilon}\right) - (1 - d_t)\frac{\phi_t}{1 - \Phi_t} \tag{2.3}$$

$$e_t^{(2)}(\beta, \eta) = d_t \left[\left(\frac{y_t - w_t\delta}{\sigma_\varepsilon}\right)^2 - 1\right] + (1 - d_t)\left(\frac{w_t\delta}{\sigma_\varepsilon}\right)\frac{\phi_t}{1 - \Phi_t} \tag{2.4}$$

where $w_t\delta = x_t\beta + v_t\alpha$, ϕ_t and Φ_t are, respectively, the normal density and cumulative distribution functions evaluated at $\frac{w_t\delta}{\sigma_\varepsilon}$. Under the true data generating process, the following is observed:[3]

$$\mathbb{E}\left[e_t^{(1)}(\beta_0, \eta_0)|x_t, z_t\right] = \mathbb{E}\left[e_t^{(2)}(\beta_0, \eta_0)|x_t, z_t\right] = 0$$

The score functions are, in matrix notation:

$$\nabla_\beta \ell_T(\beta, \eta) = \frac{1}{\sigma_\varepsilon}\, e_1' X \tag{2.5}$$

$$\nabla_\eta \ell_T(\beta, \eta) = \begin{cases} \dfrac{1}{\sigma_\varepsilon}\, e_1' V \\ \dfrac{1}{2\sigma_\varepsilon^2}\, e_2' l \\ \left[\text{vec}\left(Z'V\left(\frac{V'V}{T}\right)^{-1}\right)\right]' - \left(\alpha' \otimes \dfrac{e_1' Z}{\sigma_\varepsilon}\right) \end{cases} \tag{2.6}$$

where l is column vector whose elements are 1. The K-test is based on the asymptotic independence between the moment conditions and the expected Jacobian. Since β lies in a subset of the

[2]The pseudo-residuals play the same role as the generalized residuals proposed by Gorieroux et al. (1987)

[3]See appendix for details.

parameter space, the K-test requires the estimation of η under the null hypothesis $H_0 : \beta = \beta_0$. The identification condition for η demands that

$$\mathbb{E}\left[\nabla_{\eta\eta}\ell_t(\beta_0, \eta)\right]$$

should be continuous with respect to the parameters of the model and full rank at the true value (β_0, η_0) (see Kleibergen (2005) assumption 3). In this example, checking if the identification assumption holds is not straightforward. A practical solution is to assume identification of η.

Let $\hat{\eta}_{\beta_0}$ be a consistent estimator for η. The subsequent calculation of the K-test requires an estimator for the covariance matrix between the score and the Hessian under the null hypothesis, i.e, a statistic for:

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left\{\left[\operatorname{vec}\begin{pmatrix}\nabla_{\beta\beta}\ell_t(\beta,\eta) & \nabla_{\beta\eta}\ell_t(\beta,\eta)\\ \nabla_{\eta\beta}\ell_t(\beta,\eta) & \nabla_{\eta\eta}\ell_t(\beta,\eta)\end{pmatrix}\right]\left[\nabla_{\beta}\ell_t(\beta,\eta) \quad \nabla_{\eta}\ell_t(\beta,\eta)\right]\right\}\Bigg|_{(\beta,\eta)=(\beta_0,\hat{\eta}_{\beta_0})}$$

Finding the above covariance matrix is analytically difficult and numerical approximations can be computationally unstable in some regions of the parameter space. Thus, the use of K-test for the construction of confidence intervals are challenging.[4] In order to avoid the inherent difficulties behind the K-test I devise alternative weak instruments robust tests for censored models which are elaborated in section 3. Next subsection provides some theoretical results necessary for the derivation of the new tests.

2.3 Unrestricted Model and Its Likelihood

Instead of working directly with the structural model, I use the minimum distance framework to derive weak instruments robust tests. I first present a consistent asymptotically normal estimator for the unrestricted parameters and their covariance matrix.

From (2.2), the unrestricted conditional model is:

$$\begin{cases} Y_t = \max\{0, Z_t\pi_z + V_t\gamma + \varepsilon_t\} \\ X_t = Z_t\Pi_z + V_t \\ D_t = 1(Y_t > 0) \end{cases} \tag{2.7}$$

Under assumption 1, a simple parametric estimator for the unrestricted parameters is the TSCML. I choose the TSCML because it allows the implementation of the robust tests in almost any statistical software. Moreover, the TSCML is the same as the maximum likelihood estimator

[4]Zivot et al. (1998) have an explanation about deriving a confidence interval/set from a statistical test.

and therefore shares the efficient properties of the latter (see Newey (1987), proposition 7). In the first stage, I obtain an estimate of Π_z using the ordinary least squares. The remaining parameters are estimated from the conditional Tobit likelihood:

$$L_T^c(y_t|z_t;\pi_z,\gamma,\sigma_\varepsilon^2,\mathrm{vec}(\hat{\Pi}_z)) = \prod_{t=1}^{T}\left[\Phi\left(-\frac{z_t\pi_z+\hat{v}_t\gamma}{\sigma_\varepsilon}\right)\right]^{1-d_t}\left[\frac{1}{\sigma_\varepsilon}\,\phi\left(\frac{y_t-(z_t\pi_z+\hat{v}_t\gamma)}{\sigma_\varepsilon}\right)\right]^{d_t}$$

where $\hat{v}_t$ is the ordinary least squares residual. Instead of relying on the normality assumption, any semi-parametric estimator of the unrestricted parameters are also suitable. Some examples are the symmetrically censored least squares and the winsorized mean (see Powell (1986) and Lee (1995), respectively).

The reduced form parameters from (2.7) are identified under mild assumptions, independently of weak instruments. Moreover, the likelihood function is twice continuous differentiable. Therefore the Law of Large Numbers and the Central Limit Theorem hold under the true data generating process and the estimator for the unrestricted parameters is consistent and asymptotically normal:

Lemma 1. *If assumption 1 holds, $\mathbb{E}\|Z_t'Z_t\| < +\infty$ and $\mathbb{E}\left[Z_t'Z_t\right]$ is nonsingular, we have:*
a) the TSCML estimator for the unrestricted parameters is consistent, i.e, as $T \to +\infty$,

$$\begin{pmatrix}\hat{\pi}_z' & \hat{\theta}' & \mathrm{vec}(\hat{\Pi}_z)'\end{pmatrix} \xrightarrow{p} \begin{pmatrix}\pi_{z0}' & \theta_0' & \mathrm{vec}(\Pi_{z0})'\end{pmatrix} \tag{2.8}$$

where $\theta_0 = \begin{pmatrix}\gamma_0' & \sigma_{\epsilon 0}^2\end{pmatrix}'$.
b) As $T \to +\infty$,

$$\sqrt{T}\left[(\hat{\pi}_z-\pi_{z0})' \quad (\hat{\theta}-\theta_0)' \quad \mathrm{vec}(\hat{\Pi}_z-\Pi_{z0})'\right] \xrightarrow{d} \mathcal{N}(0, G^{-1}\Omega G^{-1\prime}) \tag{2.9}$$

where

$$G = -\begin{bmatrix}\Omega_{\pi_z\pi_z} & \Omega_{\pi_z\theta} & -\gamma_0'\otimes\Omega_{\pi_z\pi_z}\\ \Omega_{\theta\pi_z} & \Omega_{\theta\theta} & -\gamma_0'\otimes\Omega_{\theta\theta}\\ 0 & 0 & I_m\otimes\mathbb{E}\left[Z_t'Z_t\right]\end{bmatrix}, \quad \Omega = \begin{bmatrix}\Omega_{\pi_z\pi_z} & \Omega_{\pi_z\theta} & 0\\ \Omega_{\theta\pi_z} & \Omega_{\theta\theta} & 0\\ 0 & 0 & \Sigma_{\mathrm{vv}}\otimes\mathbb{E}\left[Z_t'Z_t\right]\end{bmatrix}$$

and $\begin{bmatrix}\Omega_{\pi_z\pi_z} & \Omega_{\pi_z\theta}\\ \Omega_{\theta\pi_z} & \Omega_{\theta\theta}\end{bmatrix}$ is the Fisher information matrix derived from the Tobit model.

Proof. See Appendix A.2 □

The lemma 1 shows that the presence of weak instruments does not affect the consistency of TSCML estimator for the reduced form parameters as well as the consistency of the asymptotic covariance matrix estimator.

Since our interest is to test only the structural parameter β, we focus on the restriction $\pi_z = \Pi_z\beta$. Next lemma presents the joint asymptotic distribution of $\hat{\pi}_z$ and $\hat{\Pi}_z$, which is an important result for the definition of the minimum distance objective function.

Lemma 2. *Under assumptions of lemma 1, we have:*

a) The joint asymptotic distribution of $\hat{\pi}$ *and* $\hat{\Pi}$ *is*

$$\sqrt{T}\begin{pmatrix} \hat{\pi} - \pi_0 \\ \text{vec}(\hat{\Pi} - \Pi_0) \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Omega^{-1}_{\pi_z\pi_z.\theta} + \gamma_0'\Sigma_{\text{vv}}\gamma_0 \mathbb{E}\left[Z_t'Z_t\right]^{-1} & \gamma_0'\Sigma_{\text{vv}} \otimes \mathbb{E}\left[Z_t'Z_t\right]^{-1} \\ \Sigma_{\text{vv}}\gamma_0 \otimes \mathbb{E}\left[Z_t'Z_t\right]^{-1} & \Sigma_{\text{vv}} \otimes \mathbb{E}\left[Z_t'Z_t\right]^{-1} \end{bmatrix} \right) \tag{2.10}$$

where $\Omega_{\pi_z\pi_z.\theta} = \Omega_{\pi_z\pi_z} - \Omega_{\pi_z\theta}\Omega^{-1}_{\theta\theta}\Omega_{\theta\pi_z}$.

b) Let $\hat{\Omega}_{\pi_z\pi_z.\theta}$ *be an estimator for* $\Omega_{\pi_z\pi_z.\theta}$ *as defined in the appendix. One may show that, as* $T \to +\infty$,

$$\hat{\Omega}_{\pi_z\pi_z.\theta} \xrightarrow{p} \Omega_{\pi_z\pi_z.\theta}$$

Proof. See Appendix A.3. ☐

Any statistical software with the least squares and tobit functions can provide estimates for the unrestricted parameters, $\hat{\Sigma}_{\text{vv}}$ and $\hat{\Omega}^{-1}_{\pi_z\pi_z.\theta}$. Since $\hat{\gamma}$ is a consistent estimator for γ_0, getting an estimate for the asymptotic variance in (2.10) is straightforward by the "plug-in" method.

3 Weak Instruments Robust Tests for the Endogenous Tobit Model

In this section I present the weak instruments robust tests for the endogenous Tobit model. They are modified versions of existent tests and are denoted by the subscript M. Pre-multiplying (2.10) by $\begin{bmatrix} I_k & -\beta' \otimes I_k \end{bmatrix}'$ results in

$$\sqrt{T}[(\hat{\pi}_z - \hat{\Pi}_z\beta) - (\pi_{z_0} - \Pi_{z_0}\beta)] \xrightarrow{d} \mathcal{N}\left(0, \Psi_\beta\right) \tag{3.1}$$

where:

$$\Psi_\beta = \Omega^{-1}_{\pi_z\pi_z.\theta} + (\gamma_0 - \beta)'\Sigma_{\text{vv}}(\gamma_0 - \beta)\mathbb{E}\left[Z_t'Z_t\right]^{-1}$$

Rewrite π_{z_0} as follows:

$$\pi_{z_0} = \Pi_{z_0}\beta_0 + \Pi^{\perp}_{z_0}\zeta$$

where $\Pi^{\perp}_{z_0}$ is a $k \times (k-m)$ matrix, orthogonal to Π_{z_0} and ζ is a $(k-m) \times 1$ vector. A simple weak instruments robust test for the structural parameter β is derived from the quadratic form of (3.1):

$$\text{S}(\beta) = T\left\{ \left[(\hat{\pi}_z - \hat{\Pi}_z\beta) - (\Pi_{z_0}(\beta_0 - \beta) + \Pi^{\perp}_{z_0}\zeta) \right]' \Psi^{-1}_\beta \left[(\hat{\pi}_z - \hat{\Pi}_z\beta) - (\Pi_{z_0}(\beta_0 - \beta) + \Pi^{\perp}_{z_0}\zeta) \right] \right\} \tag{3.2}$$

Under the null hypothesis $H_0^S : \beta = \beta_0,\ \zeta = 0$, $\mathrm{S}(\beta)$ converges asymptotically to a χ^2-distribution with k degrees of freedom, the number of instruments, as stated in the following theorem:

Theorem 3.1. *Define*

$$\mathrm{S_M}(\beta_0) = T\left\{\left(\hat{\pi}_z - \hat{\Pi}_z\beta_0\right)' \hat{\Psi}_{\beta_0}^{-1}\left(\hat{\pi}_z - \hat{\Pi}_z\beta_0\right)\right\} \tag{3.3}$$

where $\hat{\Psi}_{\beta_0} = \hat{\Omega}^{-1}_{\pi_z\pi_z.\theta} + (\hat{\gamma}-\beta_0)'\hat{\Sigma}_{\mathrm{vv}}(\hat{\gamma}-\beta_0)\left(\frac{Z'Z}{T}\right)^{-1}$. *Under* $H_0^S : \beta = \beta_0,\ \zeta = 0$ *and hypotheses of Lemma 2, we have:*

$$\mathrm{S_M}(\beta_0) \xrightarrow{d} \chi^2(k) \quad \textit{independently of the quality of the instruments.}$$

Proof. It follows directly from Lemma 2 and equation (3.2). □

The $\mathrm{S_M}(\beta_0)$-test is the minimum distance estimator objective function for the structural parameter evaluated at the hypothesized value.[5] As explicitly shown in (3.2), the $\mathrm{S_M}$-test tests simultaneously two hypotheses: one for the parameter value and the other for location. The second hypothesis is about the overidentification restriction.

The $\mathrm{S_M}$-test may always reject the null hypothesis parameter if $\zeta \neq 0$ and, consequently, the confidence regions constructed by inverting this statistic may be empty. On the other hand, if the instruments are weak, the test may not reject the null at any point in the parameter space, resulting in unbounded confidence regions.

In the context of linear limited dependent variable model, Anderson and Rubin (1949) proposed the following F-test:

$$\mathrm{AR}(\beta_0) = \frac{1}{k}\frac{(y - X\beta_0)'\mathrm{P}_z(y - X\beta_0)}{\hat{\sigma}^2_\varepsilon(\beta_0)} = \frac{T}{k}\left\{(\hat{\pi}_z - \hat{\Pi}_z\beta_0)'\hat{V}_{\beta_0}^{-1}(\hat{\pi}_z - \hat{\Pi}_z\beta_0)\right\} \tag{3.4}$$

where:

$$\begin{aligned} \hat{\pi}_z &= (Z'Z)^{-1}Z'y & \hat{V}_{\beta_0} &= \hat{\sigma}^2_\varepsilon(\beta_0)\left(\tfrac{Z'Z}{T}\right)^{-1} \\ \hat{\Pi}_z &= (Z'Z)^{-1}Z'X & \hat{\sigma}^2_\varepsilon(\beta_0) &= \tfrac{(y-X\beta_0)'M_z(y-X\beta_0)}{T-k} \end{aligned} \tag{3.5}$$

Therefore the $\mathrm{S_M}$-test is the extension of the AR-test to the endogenous Tobit model. Both tests project the moments onto the space spanned by the instruments (or a function of them), which does not depend on the nuisance parameter.

One disadvantage of the $\mathrm{S_M}$-test is that the degrees of freedom equal the number of excluded instruments. Thence the power against the alternative hypothesis decreases as the number of

[5]I use $\mathrm{S_M}(\beta)$ as a reference of the S-test suggested by Stock and Wright (2000).

instruments increases. This weakness motivated the development of robust tests in which the degrees of freedom and the number of structural parameters are the same.

Kleibergen's solution comes from the asymptotic independence between the moment condition and its expected Jacobian (see Kleibergen (2004) and Kleibergen (2005)). I propose to derive the robust test based on the independence between $\hat{\pi}_z - \hat{\Pi}_z\beta$, the mapping between unconstrained and constrained parameters, and the Hessian of the minimum distance function (3.3). For now, assume that $\zeta = 0$.

I start from the asymptotic joint distribution of $\hat{\pi}_z - \hat{\Pi}_z\beta$ and $\hat{\Pi}_z$.

Theorem 3.2. *Given that $\zeta = 0$, under the null hypothesis $H_0^K : \beta = \beta_0$ and assumptions of Lemma 1, the asymptotic joint distribution of $\sqrt{T}(\hat{\pi}_z - \hat{\Pi}_z\beta_0)$ and $\sqrt{T}\mathrm{vec}(\hat{\Pi}_z)$ is:*

$$\sqrt{T}\begin{bmatrix} \hat{\pi}_z - \hat{\Pi}_z\beta_0 \\ \mathrm{vec}(\hat{\Pi}_z - \Pi_{z_0}) \end{bmatrix} \xrightarrow{d} \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Psi_{\beta_0} & (\gamma_0 - \beta_0)'\Sigma_{\mathrm{vv}} \otimes (\mathbb{E}\left[Z_t'Z_t\right])^{-1} \\ \Sigma_{\mathrm{vv}}(\gamma_0 - \beta_0) \otimes (\mathbb{E}\left[Z_t'Z_t\right])^{-1} & \Sigma_{\mathrm{vv}} \otimes (\mathbb{E}\left[Z_t'Z_t\right])^{-1} \end{bmatrix}\right) \tag{3.6}$$

Proof. Immediate from Lemma 2. □

The next collorary shows that the asymptotic independence between $\sqrt{T}(\hat{\pi}_z - \hat{\Pi}_z\beta_0)$ and $\sqrt{T}(\hat{\Pi}_z - \Pi_{z_0})$ is obtained by pre-multiplying (3.6) by the lower block triangular matrix:

$$\begin{bmatrix} I_k & 0 \\ -\Sigma_{\mathrm{vv}}(\gamma_0 - \beta_0) \otimes (\Psi_{\beta_0}\mathbb{E}\left[Z_t'Z_t\right])^{-1} & I_m \otimes I_k \end{bmatrix}$$

Corollary 1. *Given $\zeta = 0$, under the null hypothesis $H_0^K : \beta = \beta_0$ and assumptions of Lemma 1, we have:*

$$\sqrt{T}\begin{bmatrix} \hat{\pi}_z - \hat{\Pi}_z\beta_0 \\ \mathrm{vec}(\bar{\Pi}_{\beta_0} - \Pi_{z_0}) \end{bmatrix} \xrightarrow{d} \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Psi_{\beta_0} & 0 \\ 0 & \Xi_{\beta_0} \end{bmatrix}\right) \tag{3.7}$$

where:

$$\bar{\Pi}_{\beta_0} = \hat{\Pi}_z - (\mathbb{E}\left[Z_t'Z_t\right])^{-1}\Psi_{\beta_0}^{-1}(\hat{\pi}_z - \hat{\Pi}_z\beta_0)(\gamma_0 - \beta_0)'\Sigma_{\mathrm{vv}}$$

$$\Xi_{\beta_0} = \Sigma_{\mathrm{vv}} \otimes \mathbb{E}\left[Z_t'Z_t\right]^{-1} - \Sigma_{\mathrm{vv}}(\gamma_0 - \beta_0)(\gamma_0 - \beta_0)'\Sigma_{\mathrm{vv}} \otimes \left(\mathbb{E}\left[Z_t'Z_t\right]\Psi_{\beta_0}\mathbb{E}\left[Z_t'Z_t\right]\right)^{-1}$$

Proof. Immediate from Theorem 3.2. □

The statistic $\bar{\Pi}_{\beta_0}$, whose distribution depends on Π_{z_0}, is a random variable which is asymptotically independent of $\hat{\pi}_z - \hat{\Pi}_z\beta_0$ under the null hypothesis H_0^K. Therefore, the distribution of $\hat{\pi}_z - \hat{\Pi}_z\beta_0$ conditional on a given value of $\bar{\Pi}_{\beta_0}$ does not depend on Π_{z_0}. I use this property to derive the modified version of the K-test.

Theorem 3.3. *Define the modified version of the* K*-test as:*

$$\begin{aligned}
\mathrm{K_M}(\beta_0) &= T\left\{\left(\hat{\pi}_z - \hat{\Pi}_z\beta_0\right)'\hat{\Psi}_{\beta_0}^{-1}\,\hat{\hat{\Pi}}_{\beta_0}\left(\hat{\hat{\Pi}}_{\beta_0}'\hat{\Psi}_{\beta_0}^{-1}\hat{\hat{\Pi}}_{\beta_0}\right)^{-1}\hat{\hat{\Pi}}_{\beta_0}'\hat{\Psi}_{\beta_0}^{-1}\left(\hat{\pi}_z - \hat{\Pi}_z\beta_0\right)\right\} \\
&= T\left\{\left(\hat{\pi}_z - \hat{\Pi}_z\beta_0\right)'\hat{\Psi}_{\beta_0}^{-\frac{1}{2}'}\hat{\hat{\mathrm{P}}}_{\beta_0}\hat{\Psi}_{\beta_0}^{-\frac{1}{2}}\left(\hat{\pi}_z - \hat{\Pi}_z\beta_0\right)\right\} \qquad (3.8)
\end{aligned}$$

where:

$$\hat{\hat{\mathrm{P}}}_{\beta_0} = \hat{\Psi}_{\beta_0}^{-\frac{1}{2}}\hat{\hat{\Pi}}_{\beta_0}\left(\hat{\hat{\Pi}}_{\beta_0}'\hat{\Psi}_{\beta_0}^{-1}\hat{\hat{\Pi}}_{\beta_0}\right)^{-1}\hat{\hat{\Pi}}_{\beta_0}'\hat{\Psi}_{\beta_0}^{-\frac{1}{2}'} \qquad (3.9a)$$

$$\hat{\hat{\Pi}}_{\beta_0} = \hat{\Pi}_z - \left(\tfrac{Z'Z}{T}\right)^{-1}\hat{\Psi}_{\beta_0}^{-1}(\hat{\pi}_z - \hat{\Pi}_z\beta_0)(\hat{\gamma} - \beta_0)'\hat{\Sigma}_{\mathrm{vv}} \qquad (3.9b)$$

$$\hat{\Psi}_{\beta_0} = \hat{\Omega}_{\pi_z\pi_z.\theta}^{-1} + (\hat{\gamma} - \beta_0)'\hat{\Sigma}_{\mathrm{vv}}(\hat{\gamma} - \beta_0)\left(\tfrac{Z'Z}{T}\right)^{-1} \qquad (3.9c)$$

Given $\zeta = 0$, under the null hypothesis $H_0^K : \beta = \beta_0$ and assumptions of Lemma 1, as $T \to +\infty$:

$$\mathrm{K_M}(\beta_0) \xrightarrow{d} \chi^2(m) \qquad (3.10)$$

regardless whether the instruments are strong, weak or irrelevant as in definition 1.

Proof. See appendix A.3. □

The negative expected value of the Hessian of the minimum distance estimator is $\Psi_{\beta_0}^{-1}\Pi_{z_0}$. Equation (3.8) shows that the $\mathrm{K_M}$-test is a quadratic form of the restriction mapping $\hat{\pi}_z - \hat{\Pi}_z\beta$ projected onto the space spanned by an estimator of the Hessian. As explained by Moreira (2003), if the Hessian is estimated independently from the restriction function, then the conditional null distribution of the test is free from the nuisance parameter.

The K-test for the linear limited dependent variable model is (see Kleibergen (2002)):

$$\mathrm{K}(\beta_0) = \frac{(Y - X\beta_0)'\,Z\tilde{\Pi}_{\beta_0}\left(\tilde{\Pi}_{\beta_0}'Z'Z\tilde{\Pi}_{\beta_0}\right)^{-1}\tilde{\Pi}_{\beta_0}'Z'\,(Y - X\beta_0)}{\hat{\sigma}_{\varepsilon}^2(\beta_0)}$$

where:

$$\tilde{\Pi}_{\beta_0} = (Z'Z)^{-1}Z'\left(X - \tfrac{(Y-X\beta_0)Y'\mathrm{M}_zX}{(Y-X\beta_0)'\mathrm{M}_z(Y-X\beta_0)}\right)$$

Using the relations in (3.5) and defining $\hat{\Sigma}_{\mathrm{uv}} = \frac{Y'M_zX}{T-k}$, $\tilde{\Pi}_{\beta_0}$ becomes:

$$\hat{\Pi}_z - \left(\tfrac{Z'Z}{T}\right)^{-1}\hat{V}_{\beta_0}^{-1}(\hat{\pi}_z - \hat{\Pi}_z\beta_0)\hat{\Sigma}_{\mathrm{uv}}$$

and the K-test is expressed as:

$$\begin{aligned}
\mathrm{K}(\beta_0) = &\ T\left\{\left(\hat{\pi}_z - \hat{\Pi}_z\beta_0\right)'\hat{V}_{\beta_0}^{-1}\,\tilde{\Pi}_{\beta_0}\left(\tilde{\Pi}_{\beta_0}'\hat{V}_{\beta_0}^{-1}\tilde{\Pi}_{\beta_0}\right)^{-1}\tilde{\Pi}_{\beta_0}'\hat{V}_{\beta_0}^{-1}\left(\hat{\pi}_z - \hat{\Pi}_z\beta_0\right)\right\} \\
= &\ T\left\{\left(\hat{\pi}_z - \hat{\Pi}_z\beta_0\right)'\hat{V}_{\beta_0}^{-\frac{1}{2}'}\,\tilde{\mathrm{P}}_{\beta_0}\hat{V}_{\beta_0}^{-\frac{1}{2}}\left(\hat{\pi}_z - \hat{\Pi}_z\beta_0\right)\right\} \qquad (3.11)
\end{aligned}$$

where:

$$\tilde{\mathrm{P}}_{\beta_0} = \hat{V}_{\beta_0}^{-\frac{1}{2}} \tilde{\Pi}_{\beta_0} \left(\tilde{\Pi}_{\beta_0}' \hat{V}_{\beta_0}^{-1} \tilde{\Pi}_{\beta_0} \right)^{-1} \tilde{\Pi}_{\beta_0}' \hat{V}_{\beta_0}^{-\frac{1}{2}'}$$

Thence, similarly to the AR-test, the K-test also has a minimum distance interpretation.

The first order condition of continuous updating estimator derived from the S_{M}-test is:

$$\begin{aligned} \frac{\partial \mathrm{S_M}(\beta)}{\partial \beta} &= -2 \left(\hat{\pi} z - \hat{\Pi}_z \beta \right)' \hat{\Psi}_\beta^{-1} \hat{\Pi}_z \; - \left[\left(\hat{\pi}_z - \hat{\Pi}_z \beta \right)' \hat{\Psi}_\beta^{-1} \otimes \left(\hat{\pi}_z - \hat{\Pi}_z \beta \right)' \hat{\Psi}_\beta^{-1} \right] \frac{\partial \mathrm{vec} \hat{\Psi}_\beta}{\partial \beta} \\ &= -2 \left\{ \left(\hat{\pi}_z - \hat{\Pi}_z \beta \right)' \hat{\Psi}_\beta^{-1} \hat{\hat{\Pi}}_\beta \right\} \end{aligned} \tag{3.12}$$

Therefore, as in the original K-test, the $\mathrm{K_M}$ is a score static with the $\mathrm{S_M}$-test as its objective functions. From (3.12), the minimum distance estimator:

$$\hat{\beta}_{\mathrm{MD}} = \left(\hat{\hat{\Pi}}_{\hat{\beta}_{\mathrm{MD}}}' \hat{\Psi}_{\hat{\beta}_{\mathrm{MD}}}^{-1} \hat{\Pi}_z \right)^{-1} \hat{\hat{\Pi}}_{\hat{\beta}_{\mathrm{MD}}}' \hat{\Psi}_{\hat{\beta}_{\mathrm{MD}}}^{-1} \hat{\pi}_z \tag{3.13}$$

will never reject the null hypothesis, implying that confidence regions derived from inverting the $\mathrm{K_M}$-test are not necessarily empty.

Nevertheless, the $\mathrm{K_M}$-test loses power at inflexion, local minimum and local maximum points, since they also satisfy (3.12). This failure is related to the underlining hypothesis that the overidentification restriction is valid, i.e $\zeta = 0$. Thence, a complementary test for overidentification is necessary, given that the value of the structural parameter is correct ($\beta = \beta_0$). The following test is the adaptation of the $\mathrm{J_K}$-test suggested by Kleibergen (2004) which is orthogonal to the $\mathrm{K_M}$-test:[6]

$$\mathrm{J}_{\mathrm{K_M}}(\beta_0) = T \left\{ \left(\hat{\pi}_z - \hat{\Pi}_z \beta_0 \right)' \hat{\Psi}_{\beta_0}^{-\frac{1}{2}'} \hat{\hat{M}}_{\beta_0} \hat{\Psi}_{\beta_0}^{-\frac{1}{2}} \left(\hat{\pi}_z - \hat{\Pi}_z \beta_0 \right) \right\} \tag{3.14}$$

$$\mathrm{J}_{\mathrm{K_M}}(\beta_0) \xrightarrow{d} \chi^2(k-m) \tag{3.15}$$

where $\hat{\hat{\mathrm{M}}}_{\beta_0} = I_k - \hat{\hat{\mathrm{P}}}_{\beta_0}$. Clearly, the $\mathrm{J}_{\mathrm{K_M}}$- and the $\mathrm{K_M}$-tests are independent. From (3.3) and (3.8) the S_{M}-test can be decomposed into two orthogonal statistics, i.e,

$$\mathrm{S_M}(\beta_0) = \mathrm{K_M}(\beta_0) + \mathrm{J}_{\mathrm{K_M}}(\beta_0)$$

At points where $\mathrm{K_M}$-test suffers spurious decline of power, $\mathrm{J}_{\mathrm{K_M}}$-test assumes the value of the $\mathrm{S_M}$-test which has always discriminatory power in those regions of the parameter space. Combining the $\mathrm{K_M}$- and the $\mathrm{J}_{\mathrm{K_M}}$-tests defines a new statistic for the structural parameter. Let $\tau_{\mathrm{K_M}}$ and $\tau_{\mathrm{J}_{\mathrm{K_M}}}$ be the levels of significance of $\mathrm{K_M}$ and $\mathrm{J}_{\mathrm{K_M}}$, respectively. The combination test $\mathrm{K_M}\mathrm{J}_{\mathrm{K_M}}$ has a

[6]The original $\mathrm{J_K}(\beta)$ is $T \left\{ \left(\hat{\pi}_z - \hat{\Pi}_z \beta_0 \right)' \hat{V}_{\beta_0}^{-\frac{1}{2}'} \tilde{\mathrm{M}}_{\beta_0} \hat{V}_{\beta_0}^{-\frac{1}{2}} \left(\hat{\pi}_z - \hat{\Pi}_z \beta_0 \right) \right\}$, where $\tilde{\mathrm{M}}_{\beta_0} = I_k - \tilde{\mathrm{P}}_{\beta_0}$

significance level of approximately $\tau = \tau_{\mathrm{K_M}} + \tau_{\mathrm{J_{K_M}}}$.[7] Since our principal interest is to test the value of the structural parameter β, a choice for τ_{K_M} is 0.04 and for $\tau_{J_{K_M}}$ is 0.01.

In the context of linear simultaneous equation models with only one endogenous variable, Moreira (2003) shows that the conditional likelihood ratio test is written as:

$$\mathrm{CLR}(\beta_0) = \frac{1}{2}\left\{\mathrm{AR}^*(\beta_0) - \mathrm{r}(\beta_0) + \sqrt{(\mathrm{AR}^*(\beta_0) + \mathrm{r}(\beta_0))^2 - 4\mathrm{J_K}(\beta_0)\mathrm{r}(\beta_0)}\right\}$$

where $\mathrm{AR}^*(\beta_0) = k \times \mathrm{AR}(\beta_0)$ and $\mathrm{r}(\beta_0)$ is a statistic that tests $\Pi_z = 0$ under the assumption that $\beta = \beta_0$. The $\mathrm{AR}^*(\beta_0)$ statistic can be decomposed as $\mathrm{AR}^*(\beta_0) = \mathrm{K}(\beta_0) + \mathrm{J_K}(\beta_0)$.[8] For the endogenous Tobit model, the modified conditional likelihood test is:

$$\mathrm{CLR_M}(\beta_0) = \frac{1}{2}\left\{\mathrm{S_M}(\beta_0) - \mathrm{r_M}(\beta_0) + \sqrt{(\mathrm{S_M}(\beta_0) + \mathrm{r_M}(\beta_0))^2 - 4\mathrm{J_{K_M}}(\beta_0)\mathrm{r_M}(\beta)}\right\} \tag{3.16}$$

where:

$$\mathrm{r_M}(\beta_0) = T\left\{\hat{\hat{\Pi}}'_{\beta_0}\hat{\Xi}^{-1}_{\beta_0}\hat{\hat{\Pi}}_{\beta_0}\right\}$$

$$\hat{\Xi}^{-1}_{\beta_0} = \hat{\Sigma}_{\mathrm{vv}}\otimes\left(\frac{Z'Z}{T}\right)^{-1} - \ \hat{\Sigma}_{\mathrm{vv}}(\hat{\gamma} - \beta_0)(\hat{\gamma} - \beta_0)'\hat{\Sigma}_{\mathrm{vv}}\otimes\left(\frac{Z'Z}{T}\hat{\Psi}_{\beta_0}\frac{Z'Z}{T}\right)^{-1}$$

The asymptotic distribution of the $\mathrm{CLR_M}$ test is not pivotal since it depends on the value of $\mathrm{r_M}(\beta_0)$. However, it is possible to simulate the critical values for the test by generating independent values of $\chi^2(m)$ and $\chi(k-m)$, as explained by Moreira (2003).[9] The $\mathrm{CLR_M}$-test is a function of the $\mathrm{K_M}$ and the $\mathrm{J_M}$ tests. Therefore there is no spurious decline of power. The limiting behavior of the $\mathrm{CLR_M}$-test as a function of $\mathrm{r_M}$ is:

$$\mathrm{CLR_M} \longrightarrow \mathrm{S_M} \quad \text{as} \quad \mathrm{r_M} \longrightarrow 0 \quad \text{and} \quad \mathrm{CLR_M} \longrightarrow \mathrm{K_M} \quad \text{as} \quad \mathrm{r_M} \longrightarrow +\infty$$

4 Power Comparison

I investigate the rejection probability of the robust tests described in the previous section simulating their power curves. I also investigate the power of the classical likelihood ratio and

[7] Let CR_{KJ}, CR_{K} and CR_{J} be the critical regions for $\mathrm{K_M J_{K_M}}$, $\mathrm{K_M}$ and $\mathrm{J_{K_M}}$ tests. Thence:

$$\begin{aligned}\Pr(\mathrm{K_M J_{K_M}} \in CR_{\mathrm{KJ}}) &= \Pr(\{\mathrm{K_M} \in CR_{\mathrm{K}}\} \cap \{\mathrm{J_M} \in CR_{\mathrm{J}}\}) + \Pr(\{\mathrm{K_M} \in CR_{\mathrm{K}}\} \cap \{\mathrm{J_{K_M}} \notin CR_{\mathrm{J}}\}) \\ &\quad + \Pr(\{\mathrm{K_M} \notin CR_{\mathrm{K}}\} \cap \{\mathrm{J_{K_M}} \in CR_{\mathrm{J}}\}) \\ &= \tau_{\mathrm{K_M}}\tau_{\mathrm{J_{K_M}}} + \tau_{\mathrm{K_M}}(1 - \tau_{\mathrm{J_{K_M}}}) + (1 - \tau_{\mathrm{K_M}})(\tau_{J_{K_M}}) \\ &= \tau_{\mathrm{K_M}} + \tau_{\mathrm{J_{K_M}}} - \tau_{\mathrm{K_M}}\tau_{\mathrm{J_{K_M}}} \approx \tau.\end{aligned}$$

[8] $\mathrm{K}(\beta_0)$ is defined in (3.11) and $\mathrm{J_K}(\beta_0)$ is defined at footnote 6.

[9] See Kleibergen (2005) for the version of the conditional robust likelihood ratio test with two or more endogenous variables.

Wald tests, which are defined as:

$$LR(\beta_0) = 2\left(\ell(\hat{\beta},\hat{\eta}) - \ell(\beta_0,\tilde{\eta}_0)\right)$$
$$W(\beta_0) = (\hat{\beta}-\beta_0)'\hat{V}(\hat{\beta})^{-1}(\hat{\beta}-\beta_0)$$

where $\ell(\hat{\beta},\hat{\eta})$ and $\ell(\beta_0,\tilde{\eta}_0)$ are the unconstrained and constrained log-likelihood functions and $\hat{V}(\hat{\beta})$ is the variance of the unrestricted estimator of β.[10] The simulations depart from the following linear latent model:

$$\begin{cases} Y_t^* = X_t\beta + U_t \\ X_t = Z_t\Pi_z + V_t \end{cases} \qquad U_t, V_t \sim \mathrm{N}\left(0, \left[\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right]\right)$$

I consider one endogenous variable and 3 instruments which come from independent random normals with zero mean and unitary variance. The instruments are the same for all simulations. The values of the concentration parameter $\lambda_z = \frac{\Pi_z' Z' Z \Pi_z}{k\sigma_v^2}$ are 20, 10, 3, 1 and 0.01, in order to mimic very strong, strong, medium, weak and inept instruments, respectively. The correlation coefficient ρ assumes values of 0, 0.5 and 0.9 and the number of observations is 300. Table 1 summarizes the simulations.

Table 1: Simulation design

k	ρ	λ_z [a]
3	0.0 0.5 0.9	20 10 3 1 0.01

[a] $\lambda_z = \frac{\Pi_z' Z' Z \Pi_z}{k\sigma_v^2}$.

It was generate 2000 endogenous Tobit samples from the above latent model. I test the hypothesis $H_0 : \beta = 0$ for each simulation and compute the proportion of rejected tests in order to build the power curves using 5% significance level. This section reports the results in which $\rho = 0.9$, leaving the remaining to appendix B.

When the instruments are very strong (λ_z equals 20) the $\mathrm{K_M}$-, $\mathrm{K_M J_{K_M}}$- and $\mathrm{CLR_M}$-tests have the same shape as the classical tests as shown in the left-top of figure 2. Therefore, the efficiency loss

[10] I use the `fmincon` function of Matlab to maximize the restricted and unrestricted loglikelihood functions.

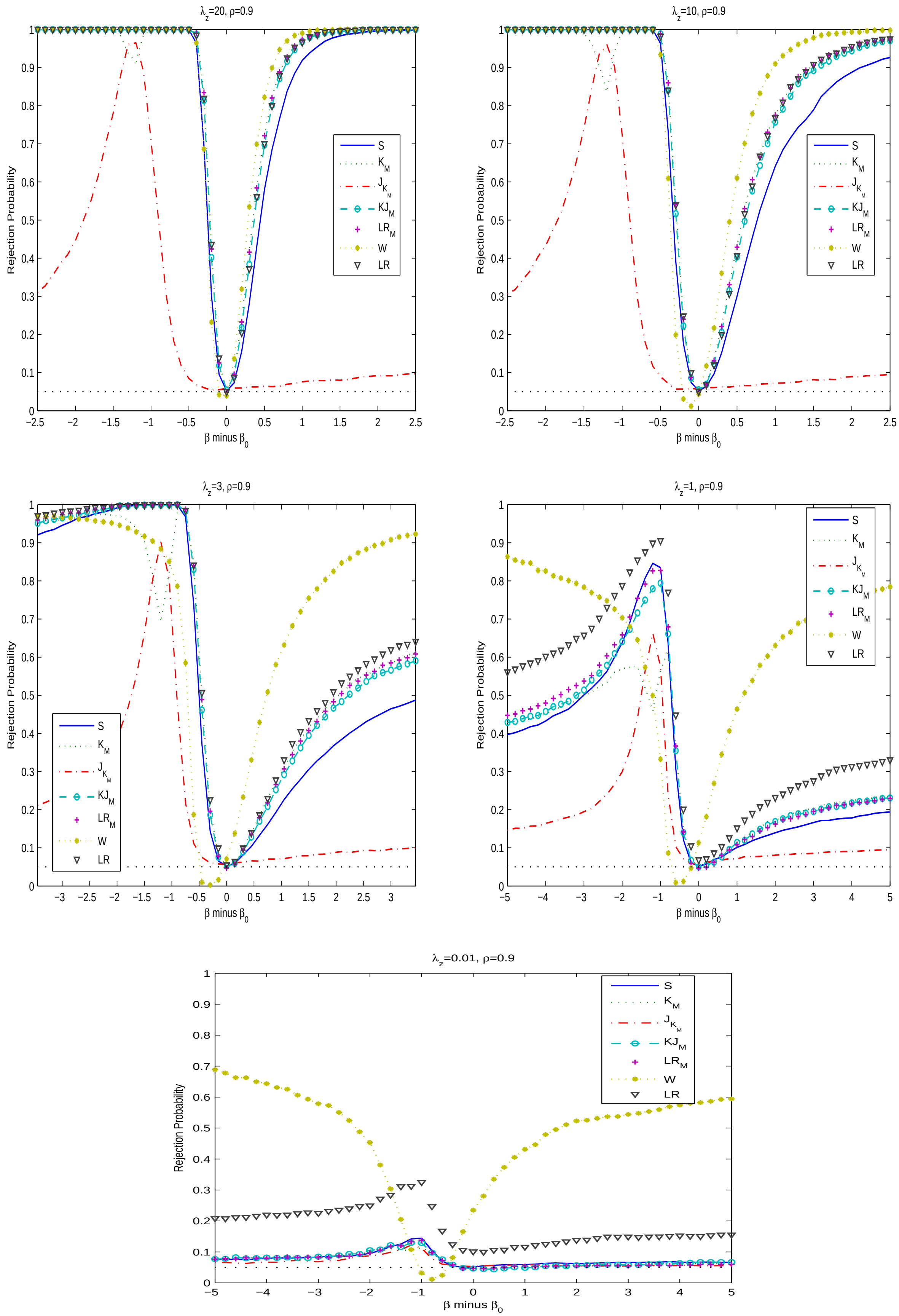
for using the former weak instruments robust tests is minimal. It is also possible to detect the gain of power of the $\mathrm{K_M}$-, $\mathrm{K_M J_{K_M}}$- and $\mathrm{CLR_M}$-tests over the $\mathrm{S_M}$-test due to model's overidentification.

As the structural parameters moves towards the lack of identification, the classical tests start to perform wrongly. When the instruments are weak, both the Wald and the LR tests underreject the null hypothesis at the true value of β (the rejection probabilities of LR and Wald are above 7% and 10%, respectively). In case of inept instruments, the rejection probability rises above 10% for the LR test and near 30% for the Wald. Since those tests are based on the unrestricted estimative of β, it is clear that the maximum likelihood estimator is biased and this bias is affecting the two classical tests.

The weak instruments robust tests perform well even when the instruments are inept, attaining the expected rejection proportion of 5% under the null hypothesis. As the identification condition fails, their power curves approximate to the horizontal line, indicating that confidence intervals derived by inverting those tests increase according to the weakness of the instruments.

Although the $\mathrm{CLR_M}$ test seems to dominate the remaining robust tests when the instruments are strong and "medium", it is not the case when the instruments are weak and inept.

Fig. 2: Power curves for Robust, Wald and LR test.



5 Empirical Application: Labor Supply of Married Females

Blundell and Walker (1986) describe a model for married female labor supply as:

$$\begin{cases} Y_t = \max\{0, X_t\beta + W_t\gamma + U_t\} \\ X_t = Z_t\Pi_z + W_t\Pi_w + V_t \\ D_t = 1(Y_t > 0) \end{cases}$$

where Y_t is weekly hours in paid work, X_t is other household income, which includes the husband's income, unearned income and savings. Besides a constant term, W_t includes demographic variables: female age and its square, education and its square, child dummy variables and a race dummy variable. The instruments Z_t include regional unemployment rate, husband occupation and housing tenure dummies. The term D_t is a labor force participation indicator. More details about the variables are in appendix B.

I use the data set from Lee (1995) obtained from the 1987 cross-section of the Michigan Panel Study of Income Dynamics. The author selected married couples with nonnegative total family income and wife not self employed at working age (18-64). 895 out of 3382, the total number of married females, were not working (approximately 26.4%). Table 2 reports estimates for the structural parameter obtained from different estimation procedures and the first-stage F-test.

Table 2: Model estimates for the structure parameter[a][b]

Method	estimate	standard deviation
maximum likelihood (mle)	0.1878	0.0612
TSCML	0.1331	0.0592
CGLS	0.1328	0.0546

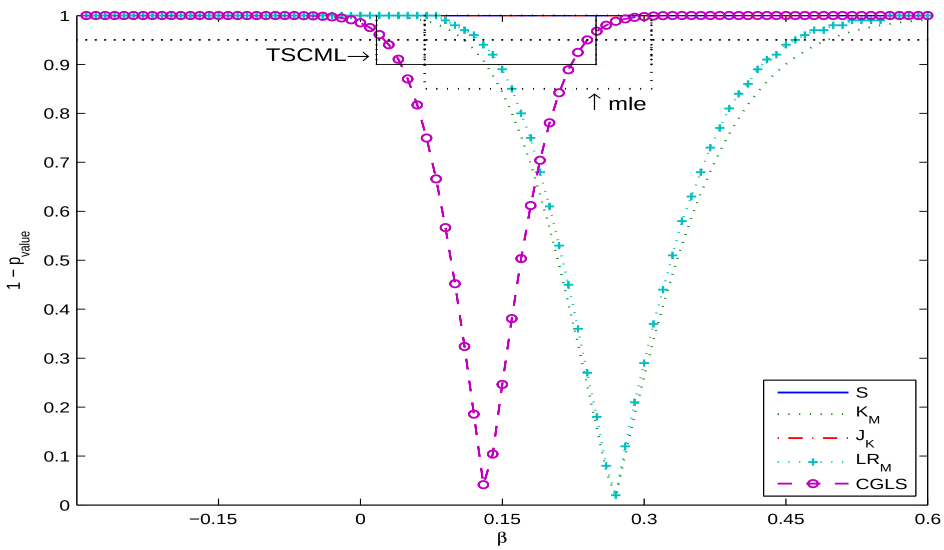
[a] 1^{st}-stage F-test = 42.23.

[b] Exogeneity $t-$test $= -5.44$

The F-test is a measure of the strength of the instruments. Since its value is above 20, it suggests that the instruments are valid. This explains why the magnitude of the estimates is almost the same. The t-test proposed by Smith and Blundell (1986) rejects the hypothesis that other income is an exogenous variable.

I use robust tests to construct 95% confidence intervals for the structural parameter β. The points of the parameter space which do not reject the hypotheses $H_0 : \beta = \beta_0$ at 5% belong to the confidence interval. The plot of the $1 - p_{\text{value}}$ function for the robust tests is shown in figure 5. The intersection between the $1 - p_{\text{value}}$ and the 95% horizontal lines delimits the confidence intervals. I also report the confidence intervals obtained by the mle, TSCML and CGLS methods.

Fig. 3: Confidence intervals for the structural parameter



In this particular example, the S_M-test generates an empty confidence interval, suggesting that the overidentification condition is not satisfied. The K_M and the CLR_M produce confidence intervals very similar to each other, but different from the ones generated by the mle, TSCLM and CGLS methods.

The graph also shows that the K_M-test is not minimized at the maximum likelihood estimator but at the minimum distance estimator (3.13). Therefore it differs from the original K-test, which attains its minimum at the limited information maximum likelihood estimator.

6 Conclusion

I show how to obtain robust tests against weak instruments for censored models with endogenous explanatory variables. These tests depart from the minimum distance objective function. This approach has two advantages: firstly it requires less restrictive assumptions about the identification of untested parameters than the K-test, and secondly, it is computationally simple to be implemented.

I carry out an empirical application of the robust test to build confidence intervals. It becomes evident that classical tests are jeopardized by the presence of weak instruments.

The proposed robust tests can be extended for other limited dependent variable models. An-other possible extension is to use semi-parametric methods for the estimation of the unrestricted parameters.

References

Amemiya, T. (1979). The estimation of simultaneous-equation tobit model. *International Economic Review 20*(1), 169–181.

Anderson, T. W. and H. Rubin (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *The Annals of Mathematical Statistics 20*(1), 46–63.

Blundell, R. and I. Walker (1986). A life-cycle consistent empirical model of family labour supply using cross-section data. *Review of Economic Studies 53*, 539–558.

Blundell, R. W. and R. J. Smith (1989). Estimation in a class of simultaneous equation limited dependent variable models. *Review of Economic Studies 56*(1), 37–58.

Gorieroux, C., A. Monfort, E. Renault, and A. Trognon (1987). Generalized residuals. *Journal of Econometrics 34*, 5–32.

Kleibergen, F. (2002, September). Pivotal statistics for testing structural parameters in instrumental variables regression. *Econometrica 70*(5), 1781–1803.

Kleibergen, F. (2004, April). Generalizing weak instrument robust iv statistics towards multiple parameters, unrestricted covariance matrices and identification statistics. Working Paper, Department of Economics, Brown University.

Kleibergen, F. (2005). Testing parameters in gmm without assuming that they are identified. *Econometrica 73*(4), 1103–1124.

Lee, L.-F. (1981). Simultaneous equations models with discrete and censored dependent variables. In C. F. Manski and D. McFadden (Eds.), *Structural Analysis of Discrete Data with Econometric Applications*, Chapter 9, pp. 346–364. MIT Press.

Lee, M.-J. (1995). Semi-parametric estimation of simultaneous equations with limited dependent variables: a case study of female labor supply. *Journal of Applied Econometrics 10*(2), 187–200.

Magnusson, L. M. (2006). Robust test against weak instruments for limited dependent variable models. Working Paper, Deparment of Economics, Brown University.

Moreira, M. J. (2003). A conditional likelihood ratio test for structural models. *Econometrica 71*(4), 1027–1048.

Newey, W. and D. McFadden (1994). Large sample estimation and hypothesis testing. In R. Engle and D. McFadden (Eds.), *Handbook of Econometrics*, Volume IV, Chapter 36, pp. 2111–2245. Elsevier.

Newey, W. K. (1987). Efficient estimation of limited dependent variable models with endogenous explanatory variables. *Journal of Econometrics* (36).

Olsen, R. J. (1978). Note on the uniqueness of the maximum likelihood estimator for the tobit model. *Econometrica 46*(5), 1211–1215.

Powell, J. L. (1986). Symmetrically trimmed least squares estimation for tobit models. *Econometrica 54*(6), 1435–1460.

Smith, R. J. and R. W. Blundell (1986). An exogeneity test for simultaneuos equation tobit model with an application to labor supply. *Econometrica 54*(3), 679–685.

Staiger, D. and J. H. Stock (1997). Instrumental variables regression with weak instruments. *Econometrica 65*(3), 557–586.

Stock, J. H. and J. Wright (2000). Gmm with weak identification. *Econometrica 68*, 1055–1096.

Zivot, E., R. Startz, and C. R. Nelson (1998). Valid confidence intervals and inference in the presence of weak instruments. *International Economic Review 39*(4), 1119–1144.

A Proofs

A.1 Some results for the Endogenous Tobit Model

A.1.1 The score as a moment restrition

The concentrated log-likelihood function for the endogenous Tobit model is :

$$\ell_T\left(\beta,\alpha,\sigma_\varepsilon^2,\Pi_z\right) \propto \sum_{t=1}^{T}(1-d_t)\ln\left(1-\Phi\left(\frac{w_t\delta}{\sigma_\varepsilon}\right)\right)+d_t\ln\left(\frac{1}{\sigma_\varepsilon}\phi\left(\frac{y_t-w_t\delta}{\sigma_\varepsilon}\right)\right) \\ -\frac{T}{2}\ln\left|\tfrac{(X-Z\Pi_z)'(X-Z\Pi_z)}{T}\right| \tag{A.1}$$

The score function is defined in (2.5) and (2.6). The expected values of $e_t^{(1)}(\beta,\eta)$ and $e_t^{(2)}(\beta,\eta)$ conditional on V_t and Z_t are:

$$\begin{aligned}
\mathbb{E}\left[e_t^{(1)}(\beta,\eta)\right] &= \mathbb{E}\left[e_t^{(1)}(\beta,\eta)\middle|\, D_t=1\right]\Pr(D_t=1)+\mathbb{E}\left[e_t^{(1)}(\beta,\eta)\middle|\, D_t=0\right]\Pr(D_t=0)\\
&= \mathbb{E}\left[\tfrac{Y_t-W_t\delta}{\sigma_\varepsilon}\middle|\, D_t=1\right]\Phi_t-\phi_t = 0\\
\mathbb{E}\left[e_t^{(2)}(\beta,\eta)\right] &= \mathbb{E}\left[e_t^{(2)}(\beta,\eta)\middle|\, D_t=1\right]\Pr(D_t=1)+\mathbb{E}\left[e_t^{(2)}(\beta,\eta)\middle|\, D_t=0\right]\Pr(D_t=0)\\
&= \phi_t\tfrac{W_t\delta}{\sigma_\varepsilon}-\left[\Phi_t-\left(1-\frac{\phi_t}{\Phi_t}\tfrac{W_t\delta}{\sigma_\varepsilon}\right)\Phi_t\right] = 0
\end{aligned}$$

A.1.2 The variance-covariance matrix for the unrestricted model

Similarly to equations (2.3) and (2.4), define the pseudo residuals:

$$\varphi_t^{(1)}(\pi_z,\theta,\Pi_z) = d_t\left(\frac{y_t-s_t\kappa}{\sigma_\varepsilon}\right)-(1-d_t)\frac{\phi_t}{1-\Phi_t} \tag{A.2}$$

$$\varphi_t^{(2)}(\pi_z,\theta,\Pi_z) = d_t\left[\left(\frac{y_t-s_t\kappa}{\sigma_\varepsilon}\right)^2-1\right]+(1-d_t)\left(\frac{s_t\kappa}{\sigma_\varepsilon}\right)\frac{\phi_t}{1-\Phi_t} \tag{A.3}$$

where $s_t\kappa = z_t\pi+v_t\gamma$, ϕ_t and Φ_t are, respectively, the normal density and cumulative distribution functions evaluated at $\frac{s_t\kappa}{\sigma_\varepsilon}$. The contribution of one observation to the score functions and to the OLS moment restriction is:

$$g_t(\pi_z,\theta,\Pi_z) = \left[\tfrac{1}{\sigma_\varepsilon}\ \varphi_t^{(1)}Z_t \quad \tfrac{1}{\sigma_\varepsilon}\ \varphi_t^{(1)}V_t \quad \tfrac{1}{2\sigma_\varepsilon^2}\ \varphi_t^{(2)} \quad \mathrm{vec}\left(X_t-Z_t\Pi_z\right)'\left(I_m\otimes Z_t\right)\right]$$

Define $\lambda(h_t)=\frac{\phi(h_t)}{\Phi(h_t)}$, the inverse Mill's ratio, evaluated at $h_t=\frac{s_t\kappa}{\sigma_\varepsilon}$. Using the information equality, the variance-covariance matrix Ω and the expected value of the Jacobian G are, respectively:

$$\Omega = \mathbb{E}\begin{bmatrix} Z_t'Z_t\iota_t^{(1)} & Z_t'V_t\iota_t^{(1)} & Z_t'\iota_t^{(2)} & 0\\ V_t'Z_t\iota_t^{(1)} & V_t'V_t\iota_t^{(1)} & V_t'\iota_t^{(2)} & 0\\ Z_t\iota_t^{(2)} & V_t\iota_t^{(2)} & \iota_t^{(3)} & 0\\ 0 & 0 & 0 & \Sigma_{\mathrm{vv}}\otimes Z_t'Z_t \end{bmatrix} \tag{A.4}$$

$$G = -\mathbb{E}\begin{bmatrix} Z_t'X_t\iota_t^{(1)} & Z_t'V_t\iota_t^{(1)} & Z_t'\iota_{2t} & -\gamma_0'\otimes Z_t'Z_t\iota_t^{(1)} \\ V_t'X_t\iota_t^{(1)} & V_t'V_t\iota_t^{(1)} & V_t'\iota_{2t} & -\gamma_0'\otimes V_t'Z_t\iota_t^{(1)} \\ X_t\iota_t^{(2)} & V_t\iota_t^{(2)} & \iota_t^{(3)} & -\gamma_0'\otimes Z_t\iota_t^{(2)} \\ 0 & 0 & 0 & I_m\otimes Z_t'Z_t \end{bmatrix} \tag{A.5}$$

where:

$$\begin{aligned}
\iota_t^{(1)} &= \frac{1}{\sigma_\varepsilon^2}\,\mathbb{E}\left[\varphi_t^{(1)}\varphi_t^{(1)}|X_t,Z_t\right] = \frac{1}{\sigma_\varepsilon^2}\left\{\Phi_t-(1-\Phi_t)\lambda'(-h_t)\right\} \\
&= \frac{1}{\sigma_\varepsilon^2}\left\{\Phi_t+\phi_t\left[\lambda(-h_t)-h_t\right]\right\} && \text{(A.6a)} \\
\iota_t^{(2)} &= \frac{1}{2\sigma_\varepsilon^3}\mathbb{E}\left[\varphi_t^{(1)}\varphi_t^{(2)}|X_t,Z_t\right] = \frac{1}{2\sigma_\varepsilon^3}\left\{(1-\Phi_t)\left(\lambda(-h_t)+\lambda'(-h_t)h_t\right)\right\} \\
&= \frac{1}{2\sigma_\varepsilon^3}\left\{\phi_t+\phi_t h_t^2-\phi_t\lambda(-h_t)h_t\right\} && \text{(A.6b)} \\
\iota_t^{(3)} &= \frac{1}{4\sigma_\varepsilon^4}\mathbb{E}\left[\varphi_t^{(2)}\varphi_t^{(2)}|X_t,Z_t\right] = \frac{1}{4\sigma_\varepsilon^4}\left\{2\Phi_t\left[1-\lambda(h_t)h_t\right]+(1-\Phi_t)\left[\lambda(-h_t)-\lambda'(-h_t)h_t\right]h_t\right\} \\
&= \frac{1}{4\sigma_\varepsilon^4}\left\{2\Phi_t+\lambda(-h_t)\phi_t h_t^2-\phi_t h_t-\phi_t^3\right\} && \text{(A.6c)}
\end{aligned}$$

One may show that $\forall h\in\mathbb{R}$, $\lambda'(-h)<0$,[11] $\lambda(-h)>h$, $1-\lambda(h)h-\lambda^2(h)>0$. Also, assuming $0<\sigma_\varepsilon<+\infty$, there exist finite real numbers such that:

$$c_1>\iota^{(1)}>0,\quad c_2>\iota^{(2)}>0\quad\text{and}\quad c_3>\iota^{(3)}>0 \tag{A.7}$$

Define $\Omega_{\pi_z\theta}=\begin{bmatrix}\Omega_{\pi_z\gamma} & \Omega_{\pi_z\sigma_\varepsilon^2}\end{bmatrix}=\Omega_{\theta\pi_z}'$ and $\Omega_{\theta\theta}=\begin{bmatrix}\Omega_{\gamma\gamma} & \Omega_{\gamma\sigma_\varepsilon^2} \\ \Omega_{\sigma_\varepsilon^2\gamma} & \Omega_{\sigma_\varepsilon^2\sigma_\varepsilon^2}\end{bmatrix}$. From (A.4) and (A.5) one may find that:

$$G^{-1}\Omega G^{-1'}=$$

$$\begin{bmatrix} \begin{bmatrix}\Omega_{\pi_z\pi_z} & \Omega_{\pi\theta} \\ \Omega_{\theta\pi_z} & \Omega_{\theta\theta}\end{bmatrix}^{-1}+\gamma'\Sigma_{vv}\gamma\begin{bmatrix}(\mathbb{E}\left[Z_t'Z_t\right])^{-1} & 0 \\ 0 & 0\end{bmatrix} & \gamma'\Sigma_{vv}\otimes\begin{bmatrix}\Omega_{\pi_z\pi_z} & \Omega_{\pi_z\theta} \\ \Omega_{\theta\pi_z} & \Omega_{\theta\theta}\end{bmatrix}^{-1}\begin{bmatrix}\Omega_{\pi_z\pi_z} \\ \Omega_{\theta\pi_z}\end{bmatrix}(\mathbb{E}\left[Z_t'Z_t\right])^{-1} \\ \Sigma_{vv}\gamma\otimes(\mathbb{E}\left[Z_t'Z_t\right])^{-1}\begin{bmatrix}\Omega_{\pi_z\pi_z} & \Omega_{\pi_z\theta}\end{bmatrix}\begin{bmatrix}\Omega_{\pi_z\pi_z} & \Omega_{\pi_z\theta} \\ \Omega_{\theta\pi_z} & \Omega_{\theta\theta}\end{bmatrix}^{-1} & \Sigma_{vv}\otimes(\mathbb{E}\left[Z_t'Z_t\right])^{-1} \end{bmatrix} \tag{A.8}$$

[11] $\lambda'(-h)=-\lambda(-h)(-h+\lambda(-h))=-\dfrac{\phi(-h)}{\Phi^2(-h)}(-h\Phi(-h)+\phi(-h))=-\dfrac{\phi(-h)}{\Phi^2(-h)}\int_{-\infty}^{-h}\Phi(w)dw<0$. The inequality holds because the integral of a strictly positive function is positive.

A.2 Proof of Lemma 1

Let $\hat{\Pi}_z$ be the ordinary least squares estimator which is derived from $Z'(X - Z\Pi_z) = 0$. Since $\mathbb{E}[V_t|Z_t] = 0$, $\mathbb{E}[Z_t' Z_t]$ is invertible and bounded, it follows that $\hat{\Pi} \xrightarrow{p} \Pi_0$. Let $\hat{\pi}_z$, $\hat{\gamma}$ and $\hat{\sigma}_\varepsilon^2$ be the solution for the conditional maximum likelihood problem evaluated at $\hat{\Pi}_z$:

$$\max_{\pi_z,\gamma,\sigma_\varepsilon^2} \frac{1}{T}\,\tilde{\ell}_T^c(\pi_z,\gamma,\sigma_\varepsilon^2) \equiv$$

$$\max_{\pi_z,\gamma,\sigma_\varepsilon^2} \frac{1}{T}\sum_{t=1}^{T}(1-d_t)\ln\left(\Phi\left(-\frac{z_t\pi_z + (x_t - z_t\hat{\Pi}_z)\gamma}{\sigma_\varepsilon^2}\right)\right) + d_t\ln\left(\frac{1}{\sigma_\varepsilon}\phi\left(\frac{y_t - z_t\pi_z - (x_t - z_t\hat{\Pi}_z)\gamma}{\sigma_\varepsilon}\right)\right) \quad \text{(A.9)}$$

Olsen (1978) proved that the log-likelihood is concave under the reparametrization $\xi_1 = \frac{\pi_z}{\sigma_\varepsilon}$, $\xi_2 = \frac{\gamma}{\sigma_\varepsilon}$ and $\xi_3 = \frac{1}{\sigma_\varepsilon}$. Since the mapping $(\xi_1, \xi_2, \xi_3) \to (\pi_z, \gamma, \sigma_\varepsilon)$ is bijective and differentiable, if $\hat{\xi} \xrightarrow{p} \psi_0$, where $\hat{\xi} = \arg\max \ell_T^c(\psi, \hat{\Pi}_z)$, then $(\hat{\pi}_z, \hat{\gamma}, \hat{\sigma_\varepsilon}) \xrightarrow{p} (\pi_{z_0}, \gamma_0, \sigma_{\epsilon 0}^2)$.

Let $\mathcal{N}_{\xi_0}$ and $\mathcal{M}_{\xi_0}$ be two open neighborhoods of ξ_0 on $\mathbb{X}$, an open convex set, such that $\mathcal{N}_{\xi_0} \subset \mathcal{M}_{\xi_0}$. I want to prove that, as $T \to \infty$, then $P(\hat{\xi} \in \mathcal{N}_{\xi_0}) \to 1$.

From concavity of the likelihood function on ξ, given that Π_z is fixed, $\ell_T(\xi, \Pi_z) \xrightarrow{p} \ell_0(\xi, \Pi_z)$ uniformly in any compact subset of $\mathbb{X}$. The limiting function $\ell_0(\xi, \Pi_z)$ is also a concave function on ξ.[12] Assume that ℓ_0 is uniquely maximized at (ξ_0, Π_{z_0}) and define the compact set $\mathcal{A} = \mathcal{N}_{\xi_0}^c \cap \mathcal{M}_{\xi_0}^c$. By the continuity of ℓ_0 there exists a ξ^* which solves $\max_{\xi\in\mathcal{A}} \ell_0(\xi, \Pi_{z_0})$.

Define $\mathcal{B} = \mathcal{A}^c \cap \mathcal{M}_{\xi_0}^c$. I claim that $\nexists\, \xi \in \mathcal{B}$ such that $\ell_0(\xi, \Pi_{z_0}) > \ell_0(\xi^*, \Pi_{z_0})$. By contradiction, suppose that $\exists \tilde{\xi} \in \mathcal{B}$ with $\ell_0(\tilde{\xi}, \Pi_{z_0}) \geq \ell_0(\xi^*, \Pi_{z_0})$. There is a line connecting ξ_0 to $\tilde{\xi}$, $\varsigma \in (0,1)$ and $\xi' \in \mathcal{A}$ such that $\xi' = \varsigma\tilde{\xi} + (1-\varsigma)\xi_0$. By concavity $\ell(\xi', \Pi_{z_0}) \geq \varsigma\ell(\tilde{\xi}, \Pi_{z_0}) + (1-\varsigma)\ell(\xi_0, \Pi_{z_0}) > \ell(\xi^*, \Pi_{z_0})$.

Set $e = \ell(\xi_0, \Pi_{z_0}) - \ell(\xi^*, \Pi_{z_0})$ and define:

$$\mathcal{C}_T^1 = \left\{\omega : |\ell_T(\xi_0, \hat{\Pi}_z) - \ell_T(\xi_0, \Pi_{z_0})| < \frac{e}{4}\right\} \qquad \mathcal{C}_T^2 = \left\{\omega : |\ell_T(\hat{\xi}, \hat{\Pi}_z) - \ell_T(\hat{\xi}, \Pi_{z_0})| < \frac{e}{4}\right\}$$
$$\mathcal{C}_T^3 = \left\{\omega : |\ell_T(\hat{\xi}, \Pi_{z_0}) - \ell_0(\hat{\xi}, \Pi_{z_0})| < \frac{e}{4}\right\} \qquad \mathcal{C}_T^4 = \left\{\omega : |\ell_T(\xi_0, \Pi_{z_0}) - \ell_0(\xi_0, \Pi_{z_0})| < \frac{e}{4}\right\}$$

Using the above inequalities one may find that: $e > \ell_0(\hat{\xi}, \Pi_{z_0}) - \ell_0(\xi_0, \Pi_{z_0})$. Thence $\bigcap_i \mathcal{C}_T^i$ implies $\hat{\xi} \in \mathcal{N}_{\xi_0}$ or $P(\bigcap_i \mathcal{C}_T^i) \leq P(\hat{\xi} \in \mathcal{N}_{\xi_0})$. Finally, one may show $P(\bigcap_i \mathcal{C}_T^i) \to 1$ as $T \to \infty$.

The asymptotic normality of the TSCML estimator follows directly from Newey and McFadden (1994) theorem 3.1 with variance covariance matrix given by:

$$G^{-1}\Omega G^{-1'} \quad \text{(A.10)}$$

where $G^{-1}\Omega G^{-1'}$ is defined in (A.8).

[12] See Newey and McFadden (1994) theorem 2.7.

A.3 Proof of Lemma 2

Define the selection matrix F as:

$$F = \begin{bmatrix} I_k & 0 & 0 \\ 0 & 0 & I_{km} \end{bmatrix}$$

Pre and pos-multiplying $G^{-1}\Omega G^{-1'}$ by F results in:

$$\begin{bmatrix} \Omega^{-1}_{\pi_z\pi_z.\theta} + \gamma_0'\Sigma_{\text{vv}}\gamma_0(\mathbb{E}\left[Z_t'Z_t\right])^{-1} & \gamma_0'\Sigma_{\text{vv}} \otimes (\mathbb{E}\left[Z_t'Z_t\right])^{-1} \\ \Sigma_{\text{vv}}\gamma_0 \otimes (\mathbb{E}\left[Z_t'Z_t\right])^{-1} & \Sigma_{\text{vv}} \otimes (\mathbb{E}\left[Z_t'Z_t\right])^{-1} \end{bmatrix} \tag{A.11}$$

where $\Omega_{\pi_z\pi_z.\theta} = \Omega_{\pi_z\pi_z} - \Omega_{\pi_z\theta}\Omega^{-1}_{\theta\theta}\Omega_{\theta\pi_z}$. Therefore, pre-multiplying (2.9) by F gives:

$$\sqrt{T}\begin{pmatrix} \hat{\pi}_z - \pi_z \\ \text{vec}(\hat{\Pi}_z - \Pi_z) \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Omega^{-1}_{\pi_z\pi_z.\theta} + \gamma_0'\Sigma_{\text{vv}}\gamma_0(\mathbb{E}\left[Z_t'Z_t\right])^{-1} & \gamma_0'\Sigma_{\text{vv}} \otimes (\mathbb{E}\left[Z_t'Z_t\right])^{-1} \\ \Sigma_{\text{vv}}\gamma_0 \otimes (\mathbb{E}\left[Z_t'Z_t\right])^{-1} & \Sigma_{\text{vv}} \otimes (\mathbb{E}\left[Z_t'Z_t\right])^{-1} \end{bmatrix}\right) \tag{A.12}$$

Let $\hat{\Omega}$, an estimator for the variance, be:

$$\hat{\Omega} = \frac{1}{T}\begin{bmatrix} Z'\hat{\Lambda}^{(1)}Z & Z'\hat{\Lambda}^{(1)}\hat{V} & Z'\hat{\Lambda}^{(2)}l & 0 \\ \hat{V}'\hat{\Lambda}^{(1)}Z & \hat{V}'\hat{\Lambda}^{(1)}\hat{V} & \hat{V}'\hat{\Lambda}^{(2)}l & 0 \\ l'\hat{\Lambda}^{(2)}Z & l'\hat{\Lambda}^{(2)}\hat{V} & l'\hat{\Lambda}^{(3)}l & 0 \\ 0 & 0 & 0 & \hat{\Sigma}_{VV} \otimes Z'Z \end{bmatrix} \tag{A.13}$$

where, for $i = 1, 2, 3$, $\hat{\Lambda}^{(i)} = \text{diag}(\iota_1^{(i)}, \dots, \iota_T^{(i)})$ evaluated at the TSCML estimator, $\hat{V} = X - Z\hat{\Pi}_z$ and $\hat{\Sigma}_{VV} = \frac{\hat{V}'\hat{V}}{T-k}$. Since the TSCML estimator is consistent and $\mathbb{E}\left[\sup \|\Omega(\pi_z, \gamma, \sigma_\varepsilon, \Pi_z)\|\right] < +\infty$,[13] $\hat{\Omega} \xrightarrow{p} \Omega$, as $T \to +\infty$ by the law of large numbers combined with the continuous mapping theorem. Consequently, $\hat{\Omega}^{-1}_{\pi_z\pi_z.\theta} \xrightarrow{p} \Omega^{-1}_{\pi_z\pi_z.\theta}$.

A.4 Proof of Theorem 3.3

Define the random variables $\mathcal{Q}$ and $\mathcal{G}$, where:

$$\sqrt{T}(\hat{\pi}_z - \hat{\Pi}_z\beta_0) \xrightarrow{d} \mathcal{G}$$
$$\sqrt{T}(\hat{\hat{\Pi}}_{\beta_0} - \Pi_{z_0}) \xrightarrow{d} \mathcal{Q}$$

$\mathcal{G}$ and $\mathcal{Q}$ are independent normal distributions (see (3.7)).

The proof is divided in three cases:

(i) The instruments are strong such that $\Pi_{z_0} = C$:

[13] See (A.7).

When the instruments are strong, $\hat{\hat{\Pi}}_{\beta_0} \xrightarrow{p} C$ which implies that:

$$\begin{aligned}\hat{\hat{\Pi}}'_{\beta_0}\hat{\Psi}_{\beta_0}^{-1}\sqrt{T}(\hat{\pi}_z - \hat{\Pi}_z\beta_0) &\xrightarrow{d} C'\Psi_{\beta_0}^{-1}\mathcal{G} \\ &\xrightarrow{d} \mathcal{N}\left(0, C'\Psi_{\beta_0}^{-1}C\right)\end{aligned}$$

Hence, the limited distribution of the $\mathrm{K_M}$-test is:

$$\begin{aligned}\mathrm{K_M}(\beta_0) &\xrightarrow{d} \mathcal{G}'\Psi_{\beta_0}^{-1}C\left(C'\Psi_{\beta_0}^{-1}C\right)^{-1}C'\Psi_{\beta_0}^{-1}\mathcal{G} \\ &\xrightarrow{d} \chi^2(m)\end{aligned}$$

(ii) The instruments are weak such that $\Pi_{z_0} = \frac{C}{\sqrt{T}}$:

When the instruments are weak we have:

$$\begin{aligned}\sqrt{T}\,\hat{\hat{\Pi}}_{\beta_0} &\xrightarrow{d} \mathcal{Q} + C \\ T\,\hat{\hat{\Pi}}'_{\beta_0}\hat{\Psi}_{\beta_0}^{-1}\hat{\hat{\Pi}}_{\beta_0} &\xrightarrow{d} (\mathcal{Q}+C)'\Psi_{\beta_0}^{-1}(\mathcal{Q}+C)\end{aligned}$$

The following conditional distribution:

$$\left[(\mathcal{Q}+C)'\Psi_{\beta_0}^{-1}(\mathcal{Q}+C)\right]^{-\frac{1}{2}}(\mathcal{Q}+C)'\Psi_{\beta_0}^{-1}\mathcal{G} \mid \mathcal{Q} \tag{A.14}$$

is $\mathcal{N}(0, I_m)$, which does not depend on $\mathcal{Q}$. Therefore, the unconditional distribution also follows a $\mathcal{N}(0, I_m)$ and the limited distribution of the $\mathrm{K_M}$-test is:

$$\mathrm{K_M}(\beta_0) \xrightarrow{d} \chi^2(m) \tag{A.15}$$

(iii) The instruments are irrelevant such that $\Pi_{z_0} = 0$

In case of irrelevant instruments we have:

$$\begin{aligned}\sqrt{T}\,\hat{\hat{\Pi}}_{\beta_0} &\xrightarrow{d} \mathcal{Q} \\ T\,\hat{\hat{\Pi}}'_{\beta_0}\hat{\Psi}_{\beta_0}^{-1}\hat{\hat{\Pi}}_{\beta_0} &\xrightarrow{d} \mathcal{Q}'\Psi_{\beta_0}^{-1}\mathcal{Q}\end{aligned}$$

One may derive the conditional distribution:

$$\left[\mathcal{Q}'\Psi_{\beta_0}^{-1}\mathcal{Q}\right]^{-\frac{1}{2}}\mathcal{Q}'\Psi_{\beta_0}^{-1}\mathcal{G} \mid \mathcal{Q} \equiv \mathcal{N}(0, I_m) \tag{A.16}$$

which does not depend on $\mathcal{Q}$. Therefore, using the same argument presented above, the limited distribution of the $\mathrm{K_M}$-test is:

$$\mathrm{K_M}(\beta_0) \xrightarrow{d} \chi^2(m) \tag{A.17}$$

B Data Description

The data set was extracted from 1987 wave of Michigan Panel Study of Income Dynamics PSID. We rescale the variables in order to match the definition used by Blundell and Smith (1989)

Table 3: Definition of the variables, 3382 observations, 895 left-censored, 1987 US PSID

Variable	definition
h_f	wife working hours per weak
w_{other}	the other household's income in \$1000
a_f [a]	$\frac{\text{Age-40}}{10}$
a_f^2	$\frac{(\text{Age}-40)^2}{100}$
ed_f [b]	(education-8)
ed_f^2	$(\text{education} - 8)^2$
C1	1 for any child between ages 0 to 5 and 0 otherwise
C2	1 for any child between ages 6 to 13 and 0 otherwise
C3	1 for any child between ages 14 to 17 and 0 otherwise
Race	1 if non-white and 0 otherwise
Tenure 1	1 if home is owned by the household and 0 otherwise
Tenure 2	1 if home is on mortgage and 0 otherwise
Husband occ 1	1 if husband is manager or professional and 0 otherwise
Husband occ 2	1 if husband is sales worker or clerical or craftsman and 0 otherwise
Husband occ 3	1 if husband is farm-related worker and 0 otherwise
un	local unemployment rate in %

[a] age of the wife in years,

[b] education of wife in years

C Power Curves

Fig. 4: Power curves for Robust, Wald and LR tests.

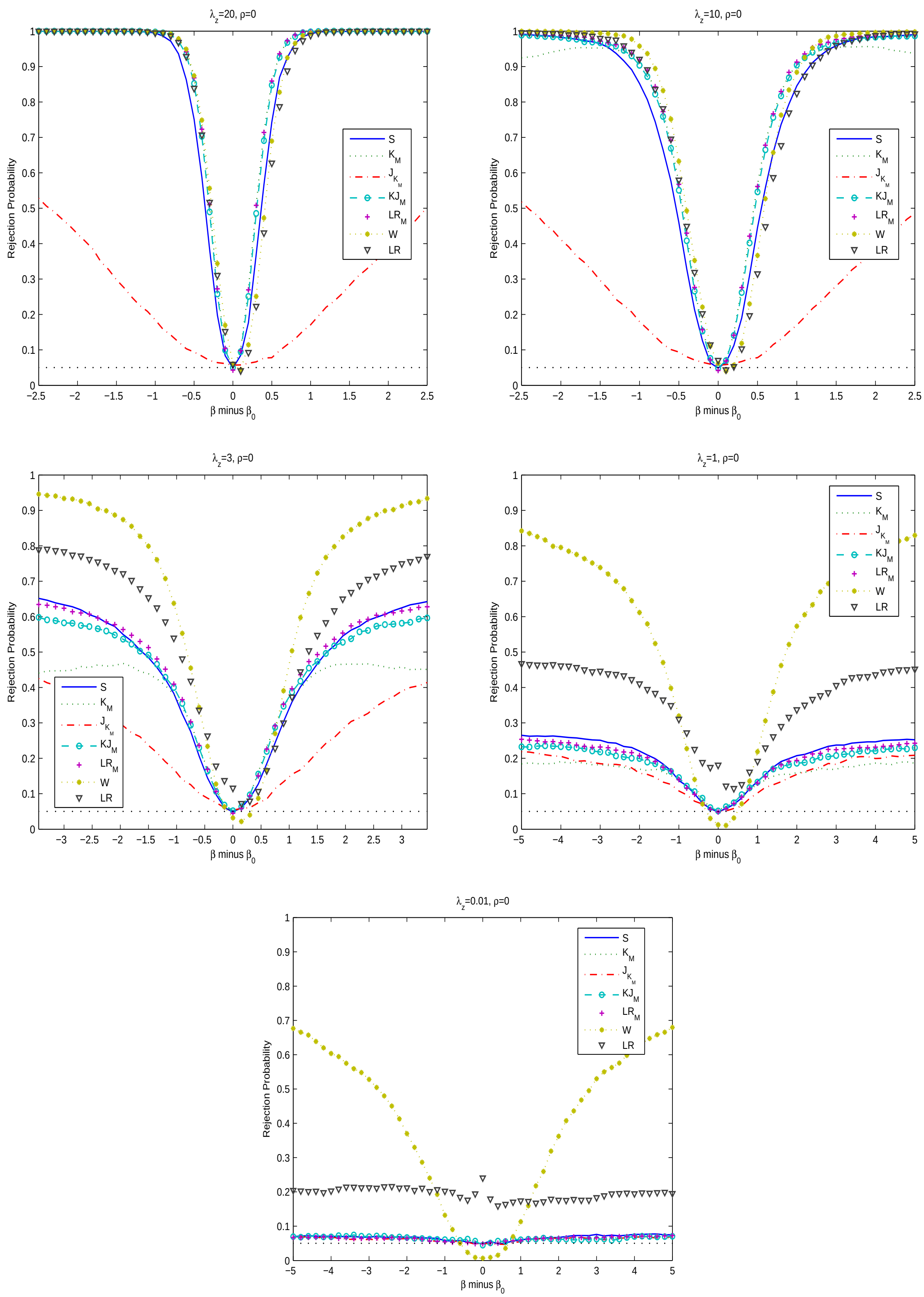


Fig. 5: Power curves for Robust, Wald and LR tests.